家蚕核型多角体病毒 *egt* 基因的分子进化分析

曹广力，贡成良[*]，薛仁宇，朱越雄，魏育红
（苏州大学医学部，江苏苏州　215123）

摘要：通过 PCR 方法获得家蚕核型多角体病毒（*Bombyx mori* nuclearpolyhedrosis virus，BmNPV）的蜕皮甾体尿苷二磷酸葡萄糖基转移酶基因（*egt*）片段，序列分析表明该片段带有 EGT 的完整 ORF，推测的多肽可形成 EGT 结构域的高级结构。为了研究 *egt* 的起源，利用家蚕基因组数据库，电子克隆了多个家蚕尿苷二磷酸葡萄糖醛酸转移酶（*UGT*）基因，在此基础上进行了进化分析，表明 BmNPV 的 EGT 为 antennal-enriched 型 UGT；推测核型多角体病毒（nucleopolyhedrivirus，NPV）和颗粒体病毒（granulovirus，GV）的 *egt* 基因在进化上来源于昆虫的 *UGT* 基因，但 GV 的 *egt* 基因在进化上的起源可能要早于 NPV 的 *egt* 基因；可能在昆虫祖先种进化形成不同昆虫目的某一时期，杆状病毒的祖先种从昆虫中获得了 antennal-enriched 型 *UGT* 基因，并进化为 *egt* 基因。家蚕的部分 *UGT* 基因与转座子元件连锁的基因组结构特点反映了杆状病毒的 *egt* 基因可能通过转座子的传递而获得。

关键词：杆状病毒；核型多角体病毒；颗粒体病毒；蜕皮甾体尿苷二磷酸葡萄糖基转移酶；尿苷二磷酸葡萄糖醛酸转移酶；进化起源

中图分类号：Q349；S881.2　　**文献标识码**：A　　**文章编号**：0454-6296（2008）12-1244-11

Analysis of molecular evolution of BmNPV *egt* gene

CAO Guang-Li，GONG Cheng-Liang[*]，XUE Ren-Yu，ZHU Yue-Xiong，WEI Yu-Hong（Medical College of Soochow University，Suzhou，Jiangsu 215123，China）

Abstract: The fragment of ecdysone UDP-glucosyltransferase gene（*egt*）was obtained by PCR from *Bombyx mori* nuclearpolyhedrosis virus（BmNPV），and sequence analysis showed that it contained the complete ORF of EGT which could form advanced structure of EGT functional region. Furthermore，for exploring the origin of *egt*，several putative uridine diphosphate glucosyl/glucuronosyltransferase genes（*UGTs*）of *B. mori* were electronically cloned. Evolutionary analysis indicated that EGT of BmNPV was of antennal-enriched UGT. Moreover，*egt* genes from nucleopolyhedrivirus（NPV）and granulovirus（GV）were presumed to originate from insect *UGT* gene，but the evolutionary origin of GV *egt* was earlier than that of NPV. During the stage of ancestor species evolving to different insects，the ancestor species of baculovirus might have obtained antennal-enriched *UGT*，which evolved to *egt* gene later. In the silkworm genome，part of *UGT* gene is closely linked with transposon，indicating that baculovirus had obtained *egt* gene through transfer of the transposon.

Key words: Baculovirus；nucleopolyhedrivirus（NPV）；granulovirus（GV）；ecdysone UDP-glucosyltransferase；uridine diphosphate glucosyl/glucuronosyltransferase（UGT）；evolutionary origin

糖基转移酶类（EC 2.4.x.y）是普遍存在于生物体内能够通过合成糖苷键将糖基连接到特定受体的一类酶，糖基供体是核苷二磷酸活化的含糖基化合物，其中 UDP-葡萄糖最为常见，另外还有 UDP-半乳糖和 UDP-鼠李糖等，供体为 UDP-葡萄糖的酶称为尿苷二磷酸葡萄糖醛酸转移酶（uridine diphosphate glucosyl/glucuronosyltransferase，UGT/UDPGT）。杆状病毒 *egt* 基因的编码产物为蜕皮甾体尿苷二磷酸葡萄糖基转移酶（ecdysone UDP-glucosyltransferase，EGT），该酶能催化昆虫体内的尿苷二磷酸葡萄糖（UDP-glucosyl）中的葡萄糖基转移到前胸腺分泌的蜕皮激素（molting hormone，MH）上，形成蜕皮激素

基金项目：国家自然科学基金项目（30671590）；高等学校博士学科点专项科研基金项目（20060285005）；江苏省教育厅科研项目（06KJB180096）；国家高技术研究发展计划项目（2006AA10A119）
作者简介：曹广力，江苏苏州人，副教授，主要从事家蚕分子遗传学研究，E-mail：guanglicao@hotmail.com
* 通讯作者 Author for correspondence，E-mail：clgong@hotmail.com
收稿日期 Received：2008-09-09；接受日期 Accepted：2008-11-26

22-O-β-D-吡喃葡萄糖，使宿主昆虫分泌的 MH 失活，从而阻止幼虫的变态，延长幼虫的取食时间，促使蛹提早成熟（Evans and O'Reilly，1998；Sakano *et al*.，2006）。*egt* 基因是一个祖先基因，包含家蚕核型多角体病毒（*Bombyx mori* nucleopolyhedrivirus，BmNPV）在内的所有杆状病毒可能都有这个基因，已知 EGT 的大小都大致相近（O'Reilly and Miller，1990）。

在动物、植物、酵母和细菌中，因有不同的 UGTs 基因被发现而被认为是一个超基因家族（Mackenzie *et al*.，1997），各亚族基因功能相似但表达有组织、器官特异性。全基因组序列研究表明，家蚕 *Bombyx mori* 中存在蜕皮激素氧化酶基因和蜕皮甾类 UDP-葡萄糖基转移酶基因，它们都参与了 MH 的代谢，对体内 MH 的水平有调节作用（Xia *et al*.，2004）。已有的研究结果显示，通过人为干预，改变昆虫的内分泌状态，可以调节昆虫的发育与变态（Dedos *et al*.，2002）。

杆状病毒 *egt* 基因位于基因组中一个缺失突变高发区内，能够在细胞继代培养中发生缺失突变，属于高度不稳定基因。删除 *egt* 基因也不影响病毒在细胞中的复制，由此推测杆状病毒 *egt* 基因可能来自宿主遗传系统（沈兴家等，2004）。由于 EGT 与 UGT 在功能上的特点，我们推测杆状病毒 *egt* 基因来自于宿主的 *UGT* 基因。为从分子水平了解 EGT 与 UGT 间的进化关系，我们通过电子克隆的方法，获得了家蚕部分的 *UGT* 基因序列，进一步通过进化分析以期了解 *egt* 基因的起源。

1　材料和方法

1.1　材料

家蚕核型多角体病毒苏州株、大肠杆菌 Top10 菌株为本实验室保存，*Taq* 酶以及限制性内切酶、T 载体购自上海生工生物工程有限公司，上下游引物均由上海生工生物工程有限公司合成，凝胶回收试剂盒购自 Biofrum 公司，T4 连接酶购自 TaKaRa 公司。

1.2　方法

1.2.1　*egt* 基因 PCR 扩增、TA 克隆及测序：BmNPV 基因组 DNA 抽提按常规方法进行。依照已公布的 BmNPV-T3 株基因组全序列（GenBank 登录号 L33180）设计 PCR 上下游引物，引物 1：5′-cgctcgagatgactattctttgctggc-3′（下划线为 *Xho* Ⅰ 酶切位点），引物 2：5′-cgggtaccgaccggcacggcaagctggcgc-3′（下划线为 *Kpn* Ⅰ 酶切位点）。PCR 扩增产物和 T 载体 pUCm-T 连接，重组质粒用 *Pst* Ⅰ 单酶切验证，序列由上海生工生物工程有限公司测定。

1.2.2　电子克隆：根据 *egt* ORF 编码的蛋白质序列和已报道的其他昆虫的 UGT 蛋白序列，利用 NCBI 在线 Blast 程序对家蚕基因组进行电子克隆，期望获得假设的家蚕 UGT 基因（*BmUGT* 基因）的 cDNA 序列。

1.2.3　进化分析：利用 NCBI 在线 Blast 可初步获得进化树，在此基础上，利用 Bioedit，MAGE 3，NTISuite 9，K-Estimator 6.1 等软件进行多重比对和进化分析。

2　结果与分析

2.1　BmNPV *egt* 基因序列分析

对克隆得到的 *egt* 基因序列（GenBank 登录号 EU350576）分析发现该序列的 ORF 为 1 521 bp。与 GenBank 中已有的 BmNPV-T3 株 *egt* 基因序列（GenBank 登录号 L33180）对比发现：该序列第 56 nt 的 C，126 nt 的 T，214 nt 的 A，216 nt 的 C，231 nt 的 T，454 nt 的 T，702 nt 的 A，714 nt 的 A，790 nt 的 C，798 nt 的 T，901 nt 的 G，946 nt 的 G，1 188 nt 的 A，1 261 nt的 A，1 288 nt 的 C，1 135 nt 的 T，在 GenBank 已有的序列中分别为 T，C，G，T，G，C，G，G，G，C，A，A，G，G，T，C。这些突变位点导致所克隆得到的 *egt* 基因的编码蛋白中，Ala^{19}，Ile^{72}，Arg^{264}，Val^{301}，Asp^{316}，Thr^{421}，Pro^{430} 在 GenBank 登录号为 L33180 序列的推导氨基酸序列中分别为 Val，Val，Gly，Ile，Asn，Asn，Ser。

2.2　BmNPV EGT 进化分析

根据家蚕 NPV 的 *egt* ORF 序列编码的蛋白质序列，利用 NCBI 在线 Blast 分析，获取核型多角体病毒（nucleopolyhedrovirus，NPV）和颗粒体病毒（granulovirus，GV）的 EGT 序列，数据来源见表 1。数据显示，带有 *egt* 基因的病毒的宿主均为鳞翅目昆虫，而在以膜翅目为宿主的松黄叶蜂 NPV（*Neodiprion sertifer* NPV，GenBank 登录号 AY430810）、冷杉锯角叶蜂（*Neodiprion abietis* NPV，GenBank 登录号 NC_008252）和红头松树叶蜂 NPV（*Neodiprion lecontei* NPV，GenBank 登录号 NC_005906）的全基因组中，未发现 *EGT* 基因。

表1 EGTs数据来源

Table 1 Source of EGT sequences

分类 Type	GenBank登录号 GenBank accession no.	种名 Species 拉丁名 Science name	种名 Species 中文名 Chinese name	缩写符号 Abbrevation
核型多角体病毒 Nucleopolyhedrovirus	NP_818769	*Adoxophyes honmai* NPV	茶小卷叶蛾 NPV	AdhoNPV
	YP_529698	*Agrotis segetum* NPV	黄地老虎 NPV	AgseNPV
	AAF15297	*Amsacta albistriga* NPV	花生黄缘灯蛾 NPV	AmalNPV
	AAK16408	*Anticarsia gemmatalis* NPV	黎豆夜蛾 NPV	AngeNPV
	NP_054044	*Autographa californica* NPV	苜蓿丫纹夜蛾 NPV	AucaNPV
	EU350576	*Bombyx mori* NPV	家蚕 NPV	BomoNPV
	AAB58353	*Buzura suppressaria* NPV	油桐尺蠖 NPV	BusuNPV
	NP_932625	*Choristoneura fumiferana* NPV	云杉卷叶蛾 NPV	ChfuNPV
	YP_249745	*Chrysodeixis chalcites* NPV	锞蚊夜蛾 NPV	ChchNPV
	YP_717663	*Clanis bilineata* NPV	豆天蛾 NPV	ClbiNPV
	YP_874315	*Ecotropis obliqua* NPV	茶尺蠖 NPV	EcobNPV
	NP_203181	*Epiphyas postvittana* NPV	苹浅褐卷蛾 NPV	EppoNPV
	NP_075195	*Helicoverpa armigera* NPV	棉铃虫 NPV	HearNPV
	NP_542753	*Helicoverpa zea* SNPV	谷实夜蛾 SNPV	HezeNPV
	YP_473326	*Hyphantria cunea* NPV	美国白蛾 NPV	HycuNPV
	YP_758442	*Leucania separata* NPV	粘虫 NPV	LeseNPV
	P41713	*Lymantria dispar* MNPV	舞毒蛾 MNPV	LydiNPV
	BAE06245	*Mamestra brassicae* MNPV	甘蓝夜蛾 MNPV	MabrNPV
	NP_689209	*Mamestra configurata* NPV	蓓带夜蛾 NPV	MacoNPV
	YP_950737	*Maruca vitrata* MNPV	豆野螟 MNPV	MaviNPV
	YP_001651040	*Orgyia leucostigma* NPV	白斑天幕毛虫 NPV	OrleNPV
	NP_046170	*Orgyia pseudotsugata* MNPV	黄杉毒蛾 MNPV	OrpsNPV
	ABE68400	*Plutella xylostella* MNPV	小菜蛾 MNPV	PlxyNPV
	NP_703005	*Rachiplusia ou* MNPV	薄荷灰夜蛾 MNPV	RaouNPV
	NP_037787	*Spodoptera exigua* NPV	甜菜夜蛾 NPV	SpexNPV
	AAP79109	*Spodoptera frugiperda* MNPV	草地夜蛾 MNPV	SpfrNPV
	2207425A	*Spodoptera littoralis* NPV	海灰翅夜蛾 NPV	SpltNPV
	ABY84898	*Spodoptera litura* NPV	斜纹夜蛾 NPV	SpluNPV
	AAZ67503	*Trichoplusia ni* SNPV	粉纹夜蛾 SNPV	TrniNPV
颗粒体病毒 Granulovirus	AAP85755	*Adoxophyes orana* GV	棉褐带卷蛾 GV	AdorGV
	YP_654536	*Choristoneura occidentalis* GV	花旗松卷蛾 GV	ChocGV
	AAX68404	*Clostera anachoreta* GV	杨扇舟蛾 GV	ClanGV
	NP_891975	*Cryptophlebia leucotreta* GV	苹果异形小卷蛾 GV	CrleGV
	NP_148925	*Cydia pomonella* GV	苹果蠹蛾 GV	CypoGV
	AAK82354	*Epinotia aporema* GV	夜小卷蛾 GV	EpapGV
	Q98166	*Lacanobia oleracea* GV	西红柿夜蛾 GV	LaolGV
	AAM70327	*Phthorimaea operculella* GV	马铃薯块茎蛾 GV	PhopGV
	AAG27416	*Plutella xylostella* GV	小菜蛾 GV	PlxyGV

SNPV：单粒包埋核型多角体病毒 Single nucleopolyhedrovirus；MNPV：多粒包埋核型多角体病毒 Multiple nucleopolyhedrovirus；GV：颗粒体病毒 Granulovirus；NPV：核型多角体病毒 Nucleopolyhedrivirus.

在对同一物种只选取1个数据的基础上，以家蚕UGT序列为外群，以Bioedit软件对蛋白序列文件编辑后，用Clustal程序对氨基酸序列进行比对，以MAGE3软件的UPGMA程序进行进化分析，结果如图1，其中BmUGT（A）和BmUGT（P）分别为家蚕的antennal-enriched型UGT（GenBank登录号NP_001040425）和phenol型UGT（GenBank登录号NP_001037040）。结果显示，EGT可分为3组，所有NPV和GV的EGT分别聚类在一起，杆状病毒的 *egt* 基因起源于antennal-enriched型 *UGT* 基因，而NPV的 *egt* 基因起源于GV的 *egt* 基因，棉褐带卷蛾 *Adoxophyes orana* GV的EGT在进化上则相对要原始一些。值得关注的是，同样以小菜蛾为宿主的PlxyGV与PlxyNPV的EGT仍然处于不同位置。利用MAGE3的"Tajima's Test of Neutrality"程序对杆状病毒EGT的中性突变检测表明，EGT多态性显著（D＝4.765498）。

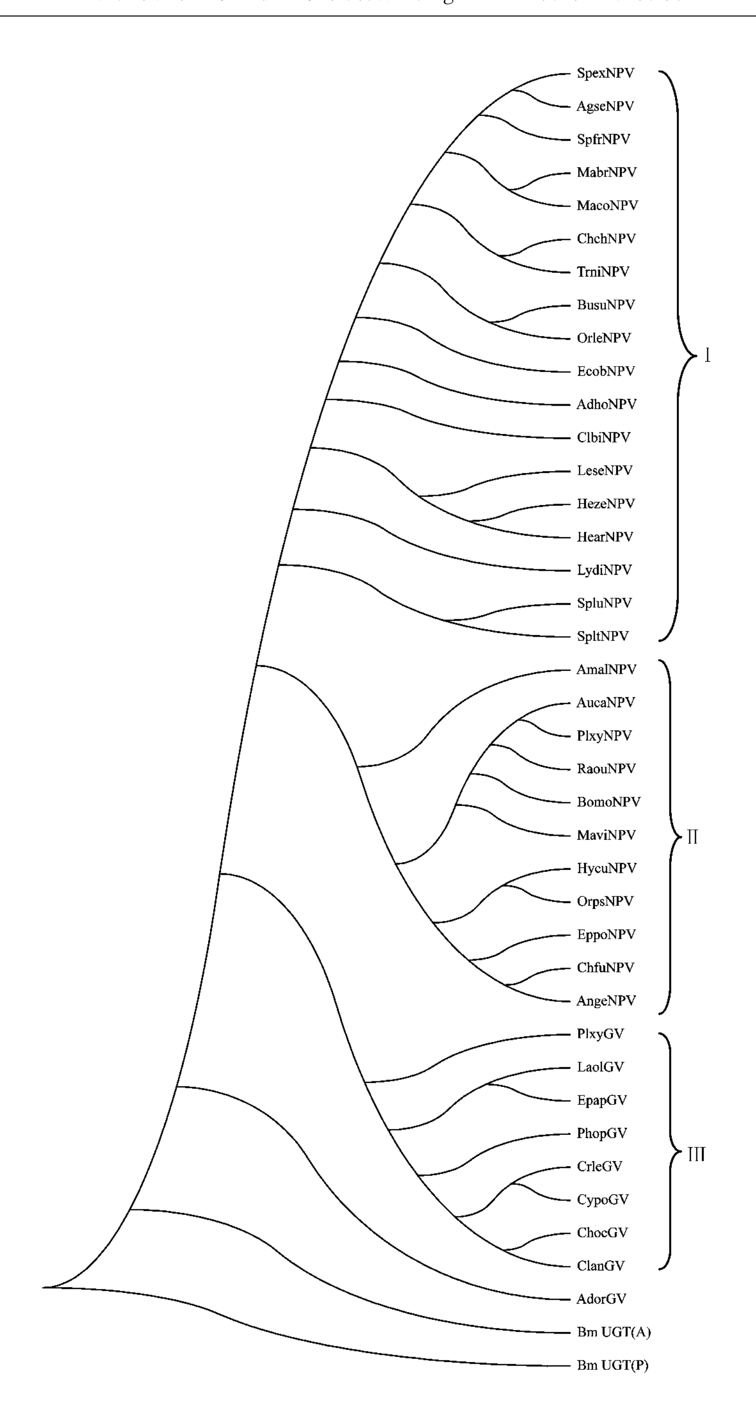


图 1　EGTs 的聚类分析

Fig. 1　Cluster analysis of EGTs

2.3　*BmUGT* 基因的预测

在昆虫中，涉及 UGTs 的蛋白质序列在 GenBank 上已有多个，包括来自于家蚕，赤拟谷盗 *Tribolium castaneum*，果蝇 *Drosophila melanogaster*，*D*. *pseudoobscura*，*D*. *willistoni*，*D*. *ananassae*，*D*. *sechellia*，*D*. *virilis*，*D*. *persimilis*，*D*. *yakuba*，冈比亚按蚊 *Anopheles gambiae*，埃及斑蚊 *Aedes aegypti*，致倦库蚊 *Culex pipiens quinquefasciatus*，豌豆蚜虫 *Acyrthosiphon pisum*，丽蝇蛹集金小蜂 *Nasonia vitripennis* 等，根据这些 UGTs 蛋白质序列通过 Blast 程序对家蚕基因组序列搜寻潜在的编码区域，并根据内含子剪接特点（GT-AG 法则），基因的 5′与 3′端非编码序列特点（5′端序列含 TATA 盒和 CAAT 盒，3′端序列含加尾信号序列 AATAAA）以及是否有 EST 序列支持等进行电子克隆，结果可获得来源于不同染色体的多条 *BmUGT* 基因的 cDNA 序列，部分电子克隆获得的

BmUGT 基因 cDNA 的 ORF 序列来源见表 2。

表 2 *BmUGTs* 电子克隆序列来源

Table 2 Source of putative *UGT* gene sequences of *Bombyx mori* obtained through the electronic clone

基因编号 Gene no.	ORF 序列来源(外显子的连接区域) Source of ORF sequences (Joining site of extrons)	EST 证据 Evidence of EST
BmUGT-1	BABH01002044(21 030 – 21 860,22 459 – 22 675,24 172 – 24 388,24 961 – 25 246)	BJ983336
BmUGT-2	BABH01002047(17 580 – 16 765,16 472 – 16 250,15 550 – 15 334,14 845 – 14 569)	CK499350
BmUGT-3	BABH01002049(6 420 – 5 602,3 667 – 3 445,2 447 – 2 231,1 280 – 962)	BP125109, CK548285
BmUGT-4	BABH01013714(400 – 600,2 600 – 3 104,3 190 – 3 413,3 850 – 4 367)	
BmUGT-5	BABH01013714(12 849 – 13 086,13 993 – 14 591,16 501 – 16 720,17 088 – 17 583)	BJ985391
BmUGT-6	BABH01029259(23 270 – 22 415,20 947 – 20 734,16 648 – 16 429,15 431 – 15 149)	DC532149, DC534966
BmUGT-7	BABH01029271(19 066 – 18 251,17 892 – 17 670,17 112 – 16 893,15 759 – 15 474)	
BmUGT-8	BABH01029271(13 903 – 13 070,11 218 – 10 996,8 832 – 8 613,7 445 – 7 157)	DN985249, CK485837
BmUGT-9	BABH01043998(8 348 – 7 527,4 720 – 4 499,3 236 – 3 020,752 – 485)	
BmUGT-10	AADK01001553(13 345 – 12 523,11 402 – 11 183,10 900 – 10 582)	
BmUGT-11	AADK01022371(4 106 – 3 273,1 421 – 1 199,641 – 418)	
BmUGT-12	BABH01029240(6 236 – 5 652,3 458 – 3 236,2 147 – 1 928,747 – 465)	
BmUGT-13	BABH01029260(2 796 – 1 878,1 553 – 1 331,932 – 730,621 – 256)	
BmUGT-14	BABH01040891(346 – 1 143,1 515-1 737,2 593 – 2 812,3 407 – 3 680)	
BmUGT-15	BABH01011830(5 731 – 5 548,4 916 – 4 703,3 680 – 3 461,3 202 <– 2 983,2 892 –> 2 663)	CK498980
BmUGT-16	BABH01011773(30 680 – 30 832), BABH01011774(2 196 – 2 283), BABH01011776(1 541 – 1 733), BABH01011778(732 – 923,1 355 – 1 571,1 840 – 2 056,3 242 – 3 461), BABH01011779(901 – 1 186)	BJ985442, AU235259
BmUGT-17	BABH01011787(13 279 – 13 133,11 860 – 11 770,10 860 – 10 665), AADK01000579(32 678 – 32 484, 32 347 – 32 125,31 257 – 31 033,30 662 – 30 406,29 814 – 29 547)	BY935347, BY914587
BmUGT-18	BABH01001663(2 642 – 2 503,1 590 – 1 407), BAAB01000846(495 – 996), BABH01042459(793 – 1 014, 1 657 – 2 219)	
BmUGT-19	BABH01011829(11 619 – 11 469,7 525 – 7 337,6 662 – 6 482,6 247 – 6 028,5 877 – 5 658,5 063 – 4 844, 2 394 – 2 120)	CK498980
BmUGT-20	BABH01019867(10 020 – 9 856,8 594 – 9 127,7 559 – 7 389,4 466 – 4 235,3 749 – 3 530,2 957 – 2 603)	BY927651, BY943958
BmUGT-21	AADK01033117(2 910 – 2 761,1 440 – 1 353), BABH01036936(1 084 – 913), BABH01036935(9 396 – 9 202,8 813 – 8 594,7 049 – 6 830,6 013 – 5 794,4 954 – 4 851,3 214 – 3 030)	DC565442, BJ984156, BY915563

其中 BABHxxxxxxxxx，AADKxxxxxxxxx 或 BAABxxxxxxxx(x 为编号)等为家蚕基因组的 GenBank 的登录号，括号中数据表明外显子所处的序列位置，而 BmUGT-15 中 3 202 <– 2 983，2 892 –> 2 663 是指 3 202 – 2 663 可能是一个外显子，也可能由于可变剪接而形成 3 202 – 2 983 和 2 892 – 2 663 的 2 个外显子，本文只使用了前一种序列进行比对。根据基因组信息分析，估计包含假基因在内的 *BmUGT* 基因有 40 个左右的成员，但由于电子克隆获得的部分 *BmUGT* cDNA 或者 ORF 较小，或者可能为假基因等原因，本文未使用而未列出。

2.4 *BmUGT* 基因的基因组信息

对电子克隆获得的 *BmUGT* 基因及基因组序列分析比较发现，*BmUGT* 基因可以以基因簇的形式紧密连锁，如 GenBank 登录号为 BABH01029271 的序列上有 3 个 *BmUGT* 基因，GenBank 登录号为 BABH01013714 的序列上有 2 个 *BmUGT* 基因。

信息分析表明，部分 *BmUGT* 基因的外显子可能因位于反转座子(retrotransposon)(Abe *et al*., 2005)中而出现在基因组的不同区域，且序列保守，如 GenBank 登录号为 BABH01029271 的序列上第 3 个 *BmUGT* 基因(参考 GenBank 登录号分别为 AADK01014331，AADK01028574，AADK01015683 的序列)，其第 2 外显子后面即为家蚕反转座子 R1Bmks 的 *gap* 基因和反转录酶(reverse transcriptase)基因；GenBank 登录号为 AADK01018081 的序列上存在 2 个连续的 *BmUGT* 基因片段，但均只有 2 个外显子，且分别与 GenBank 登录号为 BABH01029273 的序列上第 1 个和第 2 个 *BmUGT* 基因的同源性极高，这 2 个连续的 *BmUGT* 基因片段后即为反转录酶基因(图 2，未按比例)；而 GenBank 登录号为 BABH01011831的序列中，有 EST(GenBank 登录号 CK498980)支持的 *BmUGT* 基因的下游亦有 1 个反转座子；进一步分析发现，许多 *BmUGT* 基因(包括可能的 *BmUGT* 假基因)中存在有反转座子信息(甚至包含在 *BmUGT* 的外显子片段中)。

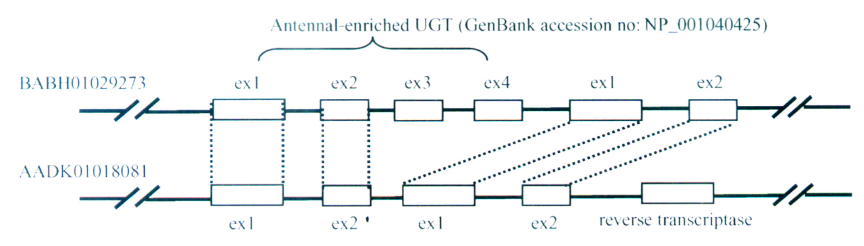


图 2　与反转录酶基因连锁的 *BmUGT* 基因片段

Fig. 2　Putative *BmUGT* gene fragments linkage with reverse transcriptase gene

信息分析还表明，部分 *BmUGT* 基因的 2 个外显子在进化上可能由 1 个外显子突变而来，如 GenBank 登录号为 BABH01011831 的序列中的外显子(1 994 – 2 527)，GenBank 登录号为 BABH01011820 的序列中的外显子(3 758 – 4 291)，GenBank 登录号为 BABH01011830 的序列中的外显子(2 673 – 3 202)，核苷酸序列非常保守，编码的氨基酸序列为 177 aa，与已报道的 UGTs 序列同源性较高，可以猜测是 1 个外显子(图 3，未按比例)，而 GenBank 登录号为 AADK01009542 的序列中的 10 552 – 10 974 区域，GenBank 登录号为 BABH01011829 的序列中的 5 658 – 6 247 区域，与上述序列在核苷酸水平上也非常保守，却包含 1 个内含子，且有 EST 序列(GenBank 登录号 CK498980)支持，可能是由于突变而在读码框中形成终止码使 1 个外显子分为 2 个外显子，在这些 *BmUGT* 基因中，虽然该区域的前后外显子也非常保守，但与该区域间的内含子长度不一，因此属于不同结构的 *BmUGT* 基因，并可能存在可变剪接而在同一编码区产生两种不同功能的 BmUGT。

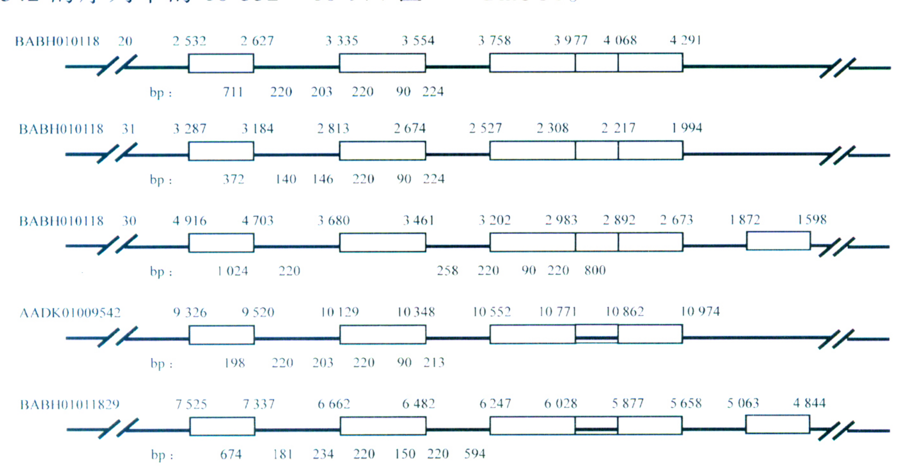
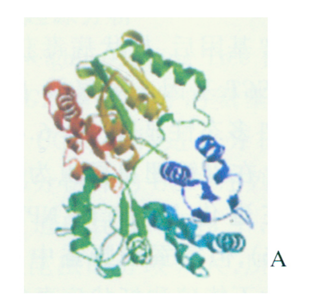


图 3　具有相同部分外显子的 *BmUGT* 基因间的比较

Fig. 3　Comparison of five putative *BmUGT* genes with homologous extrons

2.5　BmNPV EGT 和家蚕 UGT 的结构

应用 SWISS-MODEL 在线工具(http://swissmodel.expasy.org)对 BmNPV EGT 和 BmUGT 的高级结构进行预测，结果如图 4。BmNPV EGT 为本研究获得的 *egt* 基因 ORF 推测的氨基酸序列，家蚕 UGT 为随机选取的 BmUGT(表 2 的 BmUGT-3)序列。



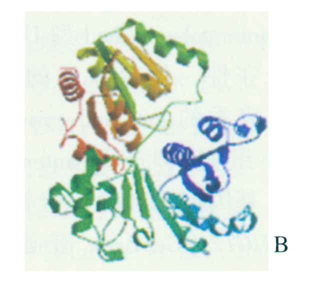


图 4　BmNPV EGT (A)和 BmUGT (B)预测的三级结构

Fig. 4　The predicated tertiary structure of BmNPV EGT (A) and putative BmUGT (B)

图中显示了 EGT 和 UGT 三级结构的高度相似性，除了 N 端和 C 端差异较明显外，其余部分几乎完全一致，这也说明了电子克隆所获得的 BmUGT 序列的可靠程度。

2.6 家蚕 *UGT* 基因的进化

对获得的所有 *BmUGT* cDNA 推导的氨基酸序列和已报道的家蚕的 UGT 序列（BmUGT-A，GenBank 登录号 NP_001040425；BmUGT-P，GenBank 登录号 NP_001037040；BmUGT2，GenBank 登录号 EU873318）进行进化分析，以苜蓿丫纹夜蛾 NPV（*Autographa californica* nucleopolyhedrovirus，AcNPV，GenBank 登录号 NP_054044）EGT 和 BmNPV EGT（本研究克隆获得，GenBank 登录号 EU350576）为外群，以 MAGE3 软件的 UPGMA 程序进行进化分析，结果如图 5 所示。

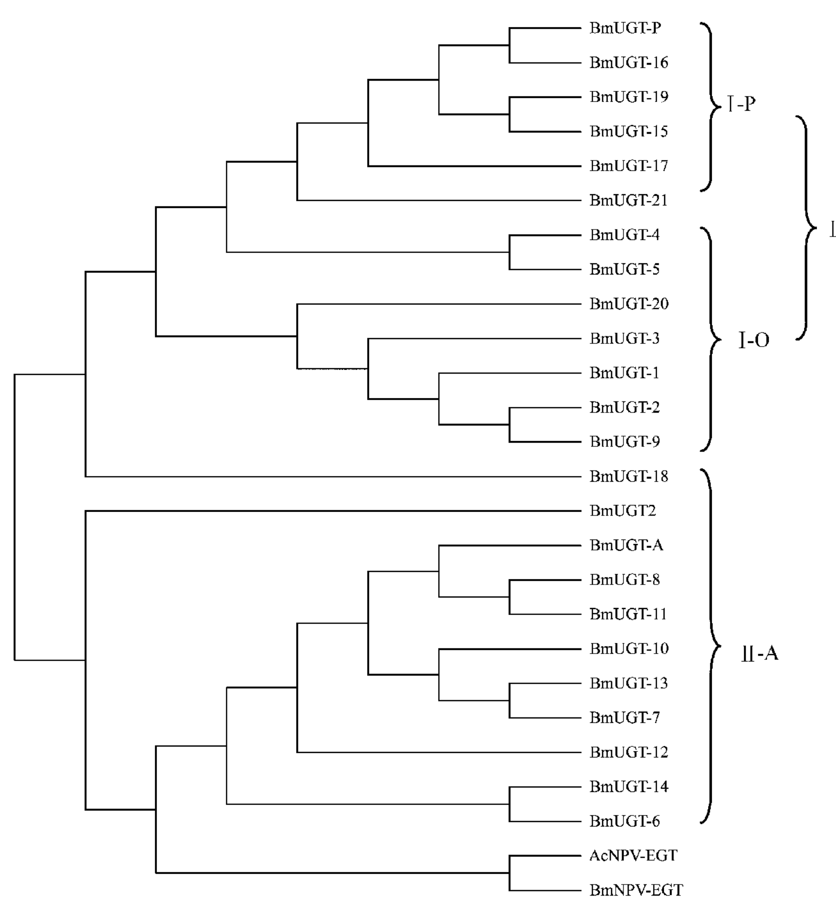


图 5 家蚕 UGTs 聚类分析

Fig. 5 Cluster analysis of UGTs of *Bombyx mori*

P：Phenol 型 UGTs Phenol UGTs；A：Antennal-enriched 型 UGTs Antennal-enriched UGTs；O：其他 UGTs Other UGTs.

结果表明家蚕 UGT（包括预测的 BmUGT）在进化上具有保守性：BmUGT 在进化上可分为两组，第 1 组分为 2 类，其中一类为 phenol 型 UGT，另一类为未定类型的 UGT；第 2 组为 antennal-enriched 型 UGT，表明了 UGT 在进化上的保守性。值得注意的是，BmNPV 和 AcNPV 的 EGT 与家蚕的 antennal-enriched 型 UGT 在进化上同处一个组群，即亦为 antennal-enriched 型，可见 NPV 的 *egt* 基因起源较晚。分析其中 GenBank 已公开的 *BmUGT2*（GenBank 登录号 EU873318）可以发现，该序列至少有 78 条 EST 支持，反映了该基因在家蚕体内是一个活性基因，从进化图可以看出，该基因可能也是家蚕的一个保守的 antennal-enriched 型 *UGT* 基因，并由此进化产生其他类型的 antennal-enriched 型 *UGT* 基因，由此推测，在祖先种昆虫的 *UGT* 基因分化为 antennal-enriched 型 *UGT* 基因后，杆状病毒获得了 UGT 的 cDNA 而进化为 *EGT* 基因。对家蚕 *UGT* 的中性突变检测表明，*UGT* 多态性显著（D = 6.462864）。

在以鳞翅目昆虫为宿主的杆状病毒中，一般认为在进化上 GV 比 NPV 更为古老（Jehle *et al*.，2006），但家蚕和野蚕中至今未发现有 GV 的存在，因此不能猜测杆状病毒的 *egt* 基因起源于家蚕或野蚕的 *UGT* 基因，但推测祖先种 GV 在进化上某一时期，其 *egt* 基因来源于宿主昆虫的 *UGT* 基因，进化为

带有 *egt* 基因的祖先种 NPV，并进一步分化为不同种的 NPV，即 GV 的 *egt* 基因在进化上的起源要早于 NPV 的 *egt* 基因。由于家蚕的部分 *UGT* 基因总是与反转座子连锁，并可能处于反转座子之中（重叠基因），且表现出高度同源性，因而对进化分析可能带来影响。由于这部分基因的 ORF 相对较小，进化分析未采用，但这些基因的存在也反映了 *UGT* 基因可以通过转座子转移到杆状病毒基因组。

2.7　对 BmNPV 的 *egt* 基因和 *BmUGT* 基因的选择压力估计

为了了解 BmNPV 的 *egt* 基因和 *BmUGT* 基因在进化过程中对编码区的选择压力估计，应用 K-Estimator 6.1 软件对部分 *egt* 和 *UGT* 基因的 DNA 序列进行了选择压力分析，其中 *egt* 基因包括来自于 AcNPV（AucaNPV，GenBank 登录号 L22858）、BmNPV（BomoNPV，GenBank 登录号 EU350576）、小菜蛾 *Plutella xylostella* 的 NPV（PlxyNPV，GenBank 登录号 DQ457003）和 GV（PlxyGV，GenBank 登录号 AF270937）、棉褐带卷蛾 GV（AdorGV，GenBank 登录号 AF547984）等杆状病毒，*UGT* 基因来自于家蚕 antennal-enriched 型 *UGT* 基因，包括 GenBank 序列（NM_001046960，EU873318）和本文电子克隆的 *BmUGT* 基因序列（*BmUGT*-6，-7，-8，-10，-11，-12，-13，-14）。选择压力分析表明（表 3），不同 *egt* 和 *UGT* 基因间的 Ka/Ks（对编码区的选择压力估计）不同，既有 Ka > Ks 的正向选择（positive or adaptive selection）基因，也有 Ka < Ks 的负向选择（negative or purifying selection）基因，如 *UGT* 基因（GenBank 登录号 EU873318）与杆状病毒的 *egt* 基因相比，主要为正向选择，与 *BmUGT* 基因相比则主要为负向选择，PlxyGV *egt* 主要为正向选择。推测 *egt* 和 *UGT* 基因在各物种间的进化既受正向选择压力作用，也有负向选择压力作用。Sliding-Windows 分析发现，基因的部分区段的 Ka/Ks < 1，少数区段表现为 Ka/Ks > 1，说明 *egt* 和 *UGT* 基因不但在不同物种间有正向与负向选择压力作用，而且在正向选择压力中也存在受负向选择压力作用的区域，反之亦然。

表 3　病毒 *egt* 基因和家蚕 *UGT* 基因的 Ka 与 Ks 平均值的比值（Ka / Ks）

Table 3　The mean Ka/Ks of *egt* gene and *UGT* gene of *Bombyx mori*

基因 Gene	编号 No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
BomoNPV egt	1														
AucaNPV *egt*	2	0.2132													
PlxyNPV *egt*	3	0.2438	0.0000												
PlxyGV *egt*	4	1.2294	1.5129	1.5651											
AdorGV *egt*	5	0.5255	0.4849	0.5130	2.9169										
NM_001046960	6	0.7785	0.6634	0.6647	0.6627	0.9955									
EU873318	7	1.3687	1.5211	1.4106	1.1279	1.0447	0.5544								
BmUGT-6	8	0.6187	0.6336	0.6080	1.4289	0.7495	0.7483	1.3418							
BmUGT-7	9	0.5539	0.4719	0.4719	0.7363	0.5496	0.5572	1.5735	0.6672						
BmUGT-8	10	0.6664	0.7633	0.7122	0.3398	0.8758	0.4343	0.7466	0.9671	1.3605					
BmUGT-10	11	1.1935	1.9532	1.5722	0.6263	1.0035	0.7126	0.9629	0.3049	0.5995	0.8051				
BmUGT-11	12	0.6801	0.7403	0.6929	0.3794	0.8513	0.4411	0.7012	0.9102	0.9398	0.1419	0.6015			
BmUGT-12	13	1.6419	1.9952	1.9488	1.4700	1.2324	1.2578	0.8375	0.9423	0.6268	0.9672	0.7575	0.9830		
BmUGT-13	14	0.7491	0.8365	0.8097	0.9125	1.1488	1.5726	0.4080	0.2901	1.1376	0.6585	0.2610	0.5325	0.5438	
BmUGT-14	15	0.7038	0.7792	0.7396	0.9178	2.6199	1.3071	0.8730	0.4067	1.4791	0.9744	0.6495	0.9149	1.2965	0.4073

2.8　EGT 与 UGT 的进化起源分析

利用 MAGE3 软件的 Neighbor-Joining Tree 程序，将电子克隆获得的家蚕 *UGT* 基因推导的氨基酸序列与 GenBank 公开的昆虫 UGTs 进行进化分析（图略，其中 GenBank 公开的部分物种的 UGTs 序列是预测的 UGT），结果发现已报道的昆虫 UGTs 中，豌豆蚜的 UGT 在进化上与其他昆虫的分歧较大，所涉及的 19 个序列聚类在一起，同样，丽蝇蛹集金小蜂中，除了 GenBank 登录号为 XP_001606484 的序列较为特殊外，其他 12 个 UGT 序列聚类在一起。来自于家蚕、埃及斑蚊、致倦库蚊、冈比亚按蚊、赤拟谷盗、果蝇等其他昆虫的 UGTs 可分为 4 个类群，而赤拟谷盗的 UGTs 可分处于 4 个类群，果蝇的 UGTs 有 2 个类群，蚊虫的 UGTs 有 2 个类群。同源性比对中，发现有保守的氨基酸残基，但未发现连续且保守的特征序列（Luque *et al*.，2002），这也许是由于 EGT/UGT 数据较丰富而显示的真实结果。

在上述 4 类 UGT 中分别选取代表种的序列，利

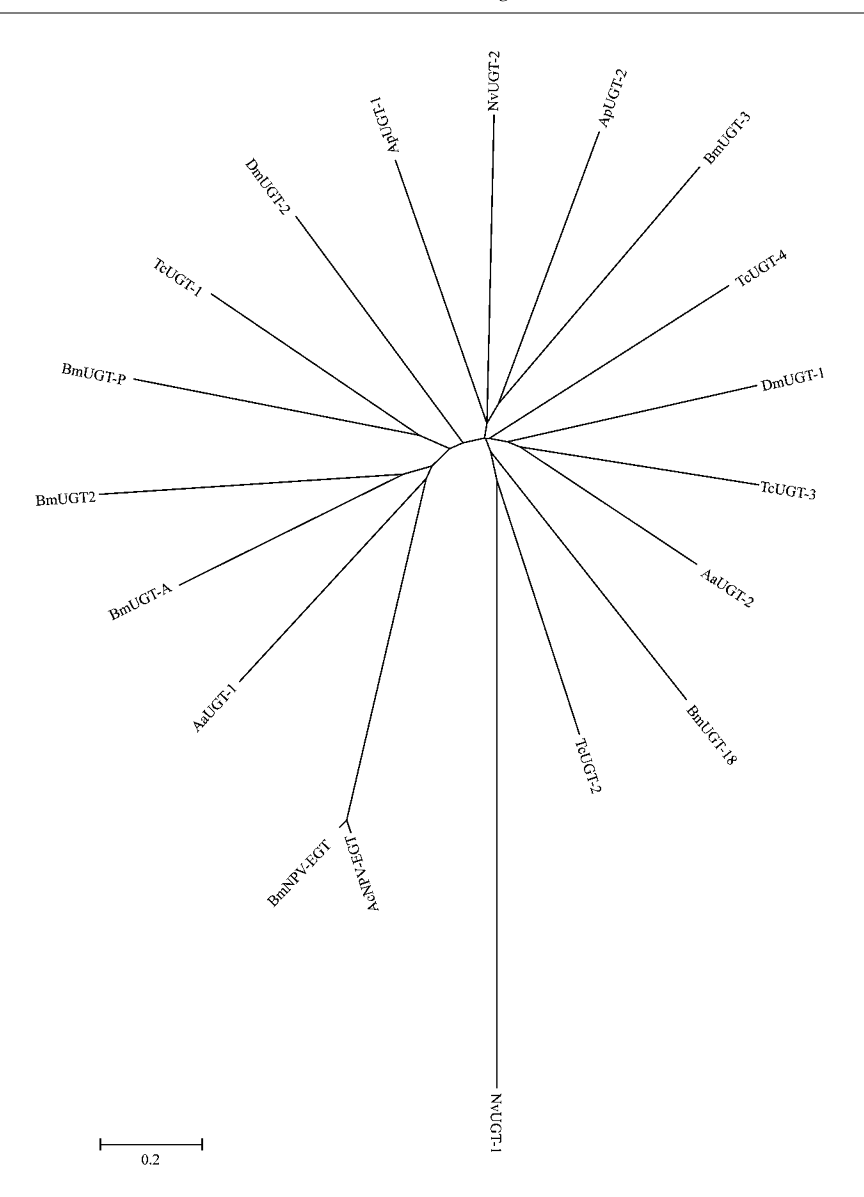


图 6 EGT 的起源分析

Fig. 6 Analysis of evolutionary origin of EGTs

用 MAGE3 软件的 Neighbor-Joining Tree 程序进行进化分析，以 AcNPV EGT（GenBank 登录号 NP_054044）和 BmNPV EGT（GenBank 登录号 EU350576）为外群，结果如图 6，其中 BmUGT-x（x 为编号，见表 2）为本研究获得的电子克隆推导序列，其他序列来源分别为：家蚕（BmUGT-A，GenBank 登录号 NP_001040425；BmUGT-P，GenBank 登录号 NP_001037040；BmUGT2，GenBank 登录号 ACF57674），埃及斑蚊（AaUGT-1，GenBank 登录号 XP_001653185；AaUGT-2，GenBank 登录号 XP_001663167），赤拟谷盗（TcUGT-1，GenBank 登录号 XP_969321；TcUGT-2，GenBank 登录号 XP_972799；TcUGT-3，GenBank 登录号XP_972444；TcUGT-4，GenBank 登录号 XP_970307），黑腹果蝇（DmUGT-1，GenBank 登录号 NP_611563；DmUGT-2：GenBank 登录号 NP_652623），豌豆蚜（ApUGT-1，GenBank 登录号 XP_001945842；ApUGT-2，GenBank 登录号 XP_001944218），丽蝇蛹集金小蜂（NvUGT-1，GenBank 登录号 XP_001606484；NvUGT-2，GenBank 登录号 XP_001603386）。图中可以看出杆状病毒的 *egt* 基因起源相对较晚，将 BmNPV EGT 通过 NCBI 的在线 Blast 分析的进化树结果与这一结果相似，在线分析结果显示了 EGT 与 BmUGT-A 最为接近，且在结果中未显示其他非昆虫物种的 UGT。从豌豆蚜[同翅目（Homoptera）]和丽蝇蛹集金小蜂[膜翅目，（Hymenoptera）]的 UGT 在进化中的聚类结果（图略）可以猜测，*egt* 基因起源于原始昆虫分化为不同物种之后的某一阶段，而根据上述昆虫所处的分类地位，推测可能在祖先种进化形成不同目的昆虫过程中，

且 *UGT* 基因已进化为不同类型的 *UGT* 基因家族后，杆状病毒的祖先种获得了 antennal-enriched 型 *UGT* 基因，并进化为 *egt* 基因。

3 讨论

研究结果表明，杆状病毒对昆虫的就眠与变态有明显影响，病毒的侵染使宿主的生长发育不再受自身的激素调节系统控制，而是更有利于病毒的复制与增殖。杆状病毒 *egt* 基因的编码产物为蜕皮甾体尿苷二磷酸葡萄糖基转移酶，该酶的代谢能使宿主昆虫分泌的 MH 失活，从而阻止幼虫的变态，延长幼虫的取食时间，促使蛹提早成熟（Evans and O'Reilly，1998）。推测认为，可能所有杆状病毒都有 *egt* 基因（闫庆生等，1999），但对松黄叶蜂 NPV、红头松树叶蜂 NPV、冷杉锯角叶蜂 NPV 和八字地老虎 GV（*Xestia c-nigrum* GV，GenBank 登录号 AF162221）的全基因组分析发现，这些 NPV 和 GV 中未发现有 *egt* 基因基因，进一步分析表明，基因组中含有不完整的 EGT 同源的编码序列，即基因组中含有 *egt* 基因的痕迹，而黄地老虎 GV（*Agrotis segetum* GV，GenBank 登录号 AY522332）中存在有编码 207 个氨基酸残基的类似 *egt* 基因序列。这些证据表明杆状病毒 *egt* 基因在进化过程中有可能发生缺失，这也许与这些病毒的宿主的生活环境、生理特点等相关。杆状病毒 *egt* 基因位于基因组中一个缺失突变高发区内，能够在细胞继代培养中发生缺失突变，属于高度不稳定基因。删除 *egt* 基因也不影响病毒在细胞中的复制，由此推测杆状病毒 *egt* 基因可能来自宿主遗传系统（沈兴家等，2004）。

对克隆的 BmNPV 的 *egt* 基因分析发现，在其 ORF 内含有多个突变位点，但可能不影响其编码蛋白 EGT 功能，EGT 高级结构预测表明其具有 UGT 结构特点。杆状病毒的 *egt* 基因进化分析表明，EGT 可分为 3 组，与宿主的共进化关系不明显；推测 GV 的 *egt* 基因起源于宿主昆虫的 antennal-enriched 型 *UGT* 基因，而 NPV 的 *egt* 基因起源于 GV 的 *egt* 基因。

根据 CAZy（http：//www.cazy.org/）数据库，糖基转移酶超基因家族依据其序列相似性和底物特异性被分为 91 个家族（截止到 2008 年 8 月），*UGT* 基因家族为超基因家族，这不但表现在 *UGT* 基因在生物群体中的多样性，而且对某一物种而言，也存在 UGT 超基因家族。在家蚕中，http：//silkworm.genomics.org.cn/网站进行的初步预测表明，*UGT* 基因有 37 个成员，我们根据基因组信息分析，估计包含假基因在内的 *UGT* 基因确实有 40 个左右的成员。电子克隆 *UGT* 基因发现基因组序列中，家蚕 *UGT* 基因可以以基因簇的形式紧密连锁。

由于 BmNPV 的 EGT 与家蚕的 antennal-enriched 型 UGT 在进化上同处一个类群，可见 NPV 的 *egt* 基因是起源于昆虫宿主的 UGT 分化为不同型 *UGT* 基因之后，并进化为 *egt* 基因。GV 的 EGT 与昆虫宿主的 UGT 在聚类上分歧更小，表明 GV 的 *egt* 基因在进化上的起源要早于 NPV 的 *egt* 基因，这与其他基因的进化研究结果相同（王崇龙等，2006）。在进化上鳞翅目昆虫是起源较晚的昆虫，虽然家蚕 NPV 的 EGT 与家蚕的 UGT 序列同源性高可以是由于共进化形成，但总体上，由于杆状病毒的 EGT（包括 NPV 和 GV 的 EGT）与家蚕的 UGT 序列的同源性，相对高于与其他昆虫的 UGT 序列的同源性，这也反映了杆状病毒的 *egt* 基因起源于昆虫的 *UGT* 基因的时期较晚。在以鳞翅目昆虫为宿主的杆状病毒在进化上 NPV 要晚于 GV，但家蚕和野蚕中至今未发现有 GV 的存在，因此还不能确定杆状病毒的 EGT 是否起源于鳞翅目昆虫宿主。随着不同物种的 UGT 数据增多，人们将有可能进一步了解 UGT 起源问题，包括昆虫 UGT 的起源。

参 考 文 献（References）

Abe H，Mita K，Yasukochi Y，Oshiki T，Shimada T，2005. Retrotransposable elements on the W chromosome of the silkworm，*Bombyx mori*. *Cytogenet*. *Genome Res*.，110：144－151.

Dedos SG，Szurdoki F，Szekacs A，Mizoguchi A，Fugo H，2002. Induction of dauer pupae by fenoxycarb in the silkworm，*Bombyx mori*. *J*. *Insect Physiol*.，48(9)：857－865.

Evans OP，O'Reilly DR，1998. Purification and kinetic analysis of a baculovirus ecdysteroid UDP-glucosyltransferase. *Biochem*. *J*.，330：1 265－1 270.

Jehle JA，Lange M，Wang H，Hu Z，Wang Y，Hauschild R，2006. Molecular identification and phylogenetic analysis of baculoviruses from Lepidoptera. *Virology*，346：180－193.

Luque T，Okano K，O'Reilly DR，2002. Characterization of a novel silkworm（*Bombyx mori*）phenol UDP-glucosyltransferase. *Eur*. *J*. *Biochem*.，269：819－825.

Mackenzie PI，Owens IS，Burchell B，Bock KW，Bairoch A，Bélanger A，Fournel-Gigleux S，Green M，Hum DW，Iyanagi T，Lancet D，Louisot P，Magdalou J，Chowdhury JR，Ritter JK，Schachter H，Tephly TR，Tipton KF，Nebert DW，1997. The UDP glycosyltransferase gene superfamily：Recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics*，7(4)：255－269.

O’Reilly DR，Miller LK，1990. Regulation of expression of a baculovirus pesticide ecdysteroid UDP-glucosyltransferase gene. *J. Virol.*，64(3)：1 321 – 1 328.

Sakano D，Li B，Xia Q，Yamamoto K，Fujii H，Aso Y，2006. Genes encoding small heat shock protein of the silkworm，*Bombyx mori*. *Biosci. Biolechnol. Biochem.*，70(10)：2 443 – 2 450.

Shen XJ，Zhang ZF，Li YR，2004. Research progress in insect baculovirus *egt* gene. *China Sericulture*，25(2)：4 – 7. [沈兴家，张志芳，李奕仁，2004. 昆虫杆状病毒 *egt* 基因研究进展. 中国蚕业，25(2)：4 – 7]

Wang CL，Xue RY，Cao GL，Shen WD，Gong CL，2006. Sequence analysis of BmNPV *alk-exo* gene and molecular evolution of baculovirus. *Science of Sericulture*，32(2)：199 – 205. [王崇龙，薛仁宇，曹广力，沈卫德，贡成良，2006. 家蚕核型多角体病毒碱性磷酸酶基因(*alk-exo*)与杆状病毒的分子进化分析. 蚕业科学，32(2)：199 – 205]

Xia Q，Zhou Z，Lu C *et al.*，2004. A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science*，306：1 937 – 1 940.

Yan QS，Ouyang XG，Li CB，Pang Y，1999. Insect baculovirus-encoded ecdysteroid UDP-glucosyltransferases. *Progress in Biochemistry and Biophysics*，26(3)：241 – 242. [闫庆生，欧阳晓光，李充璧，庞义，1999. 昆虫杆状病毒的蜕皮甾体尿苷二磷酸葡糖转移酶. 生物化学与生物物理进展，26(3)：241 – 242]

(责任编辑：赵利辉)